

WHAT IS CLAIMED IS:

SUB A1

1. A method for performing a clean operation on an input table having an
input table name, comprising:
2 receiving at least one rule definition, wherein each rule definition indicates a find
criteria, a replacement value, and an input data column in the input table;
3 searching, for each rule definition, the input data column for any fields that match
the find criteria; and
4 inserting, for each rule definition, the replacement value in the fields in the input
data column that match the find criteria, wherein subsequent applications of additional
rule definitions applied to the same input data column operate on replacement values
inserted in the input data column in previously applied rule definitions.

1 2. The method of claim 1, wherein each rule definition is associated with one
rule table including the find criteria and replacement value, wherein a rule table column
parameter for each rule definition indicates the columns in the rule table including the
find criteria and replacement value for the rule definition.

1 3. The method of claim 1, wherein there is a separate rule table including the
find criteria and replacement value associated with at least one rule definition, wherein,
for each rule definition, a rule table column parameter indicates the columns in the rule
table for the rule definition including the find criteria and replacement value for that rule
definition.

1 4. The method of claim 1, wherein the input data column for a first and
second applied rule definitions is the same input data column, wherein the replacement
value for the first rule definition is inserted into at least one field in the input data
column, and wherein the find criteria of the second rule definition is applied to the
replacement value inserted in the input data column.

1 5. The method of claim 1, wherein at least one rule definition includes
2 multiple find criteria and a corresponding replacement value for each find criteria,
3 wherein the step of searching the input data column comprises applying each of the
4 multiple find criteria to one field until one of: (i) a match occurs and (ii) none of the
5 multiple find criteria are found to match the field content, and wherein inserting the
6 replacement value comprises inserting the replacement value corresponding to one find
7 criteria that matched the field content.

1 6. The method of claim 5, wherein a sort column includes values to use to
2 sort the multiple find criteria and corresponding replacement value, wherein the step of
3 searching comprises applying the multiple find criteria to each field in the order specified
4 in the sort column.

1 7. The method of claim 1, wherein the rule definition comprises a type of
2 rule that is a member of the set of rules consisting of: find and replace, discretization, and
3 numeric clip, wherein at least two rule definitions are comprised of different rule types.

1 8. The method of claim 1, wherein the find criteria for at least one rule
2 definition comprises an upper bound and lower bound, wherein searching comprises
3 searching for any fields that have values within the upper and lower bounds.

1 9. The method of claim 8, wherein the at least one rule definition including
2 find criteria having upper and lower bounds includes multiple find criteria and a
3 corresponding replacement value for each find criteria, wherein the step of searching the
4 input data column comprises applying each of the multiple find criteria to one field until
one of: (i) a match occurs and (ii) none of the multiple find criteria are found to match the
field content, and wherein inserting the replacement value comprises inserting the
replacement value corresponding to one find criteria that matched the field content.

1 10. The method of claim 8, wherein searching comprises searching for any
2 fields that have values outside of one of the upper and lower bounds.

1 11. The method of claim 1, wherein the find criteria for at least one rule
2 definition comprises an upper bound and lower bound and wherein the replacement value
3 is an upper replacement value and further comprising a lower replacement value, wherein
4 searching comprises searching for any fields that have values within the upper and lower
5 bounds and wherein inserting comprises inserting the upper replacement value if the field
6 has a value greater than the upper bound and inserting the lower replacement value if the
7 field has a value less than the lower bound.

1 12. The method of claim 11, wherein the at least one rule definition including
2 find criteria having upper and lower bounds includes multiple find criteria and a
3 corresponding upper and lower replacement value for each find criteria, wherein the step
4 of searching the input data column comprises applying each of the multiple find criteria
5 to one field until one of: (i) a match occurs and (ii) none of the multiple find criteria are
6 found to match the field content, and wherein inserting the replacement value comprises
7 inserting the replacement value corresponding to one find criteria that matched the field
8 content.

1 13. The method of claim 1, wherein the rule definitions include a row clean
2 flag, and wherein at least one rule definition has the row clean flag set, further comprising
3 removing any row including a field matching the search criteria from the input table when
4 the row clean flag is set.

1 14. A system for performing a clean operation on an input table having an
2 input data table name, comprising;

3 means for receiving at least one rule definition, wherein each rule definition
4 indicates a find criteria, a replacement value, and an input data column in the input table;
5 means for searching, for each rule definition, the input data column for any fields
6 that match the find criteria; and
7 means for inserting, for each rule definition, the replacement value in the fields in
8 the input data column that match the find criteria, wherein subsequent applications of
9 additional rule definitions applied to the same input data column operate on replacement
10 values inserted in the input data column in previously applied rule definitions.

1 15. The system of claim 14, wherein each rule definition is associated with
2 one rule table including the find criteria and replacement value, wherein a rule table
3 column parameter for each rule definition indicates the columns in the rule table
4 including the find criteria and replacement value for the rule definition.

1 16. The system of claim 14, wherein there is a separate rule table including the
2 find criteria and replacement value associated with at least one rule definition, wherein,
3 for each rule definition, a rule table column parameter indicates the columns in the rule
4 table for the rule definition including the find criteria and replacement value for that rule
5 definition.

1 17. The system of claim 14, wherein the input data column for a first and
2 second applied rule definitions is the same input data column, further comprising:
3 means for inserting the replacement value for the first rule definition into at least
4 one field in the input data column; and
5 means for applying the find criteria of the second rule definition to the
6 replacement value inserted in the input data column.

1 18. The system of claim 14, wherein at least one rule definition includes
2 multiple find criteria and a corresponding replacement value for each find criteria,
3 wherein the means for searching the input data column comprises applying each of the
4 multiple find criteria to one field until one of: (i) a match occurs and (ii) none of the
5 multiple find criteria are found to match the field content, and wherein inserting the
6 replacement value comprises inserting the replacement value corresponding to one find
7 criteria that matched the field content.

1 19. The system of claim 18, wherein a sort column includes values to use to
2 sort the multiple find criteria and corresponding replacement value, wherein the means
3 for searching comprises applying the multiple find criteria to each field in the order
4 specified in the sort column.

1 20. The system of claim 14, wherein the rule definition comprises a type of
2 rule that is a member of the set of rules consisting of: find and replace, discretization, and
3 numeric clip, wherein at least two rule definitions are comprised of different rule types.

1 21. The system of claim 14, wherein the find criteria for at least one rule
2 definition comprises an upper bound and lower bound, wherein the means for searching
3 comprises searching for any fields that have values within the upper and lower bounds.

1 22. The system of claim 21, wherein the at least one rule definition including
2 find criteria having upper and lower bounds includes multiple find criteria and a
3 corresponding replacement value for each find criteria, wherein the means for searching
4 the input data column comprises applying each of the multiple find criteria to one field
5 until one of: (i) a match occurs and (ii) none of the multiple find criteria are found to
6 match the field content, and wherein the means for inserting the replacement value

7 comprises inserting the replacement value corresponding to one find criteria that matched
8 the field content.

1 23. The system of claim 20, wherein the means for searching comprises
2 searching for any fields that have values outside of one of the upper and lower bounds.

1 24. The system of claim 14, wherein the find criteria for at least one rule
2 definition comprises an upper bound and lower bound and wherein the replacement value
3 is an upper replacement value and further comprising a lower replacement value, wherein
4 the means for searching comprises searching for any fields that have values within the
5 upper and lower bounds and wherein inserting comprises inserting the upper replacement
6 value if the field has a value greater than the upper bound and inserting the lower
7 replacement value if the field has a value less than the lower bound.

1 25. The system of claim 24, wherein the at least one rule definition including
2 find criteria having upper and lower bounds includes multiple find criteria and a
3 corresponding upper and lower replacement value for each find criteria, wherein the
4 means for searching the input data column comprises applying each of the multiple find
5 criteria to one field until one of: (i) a match occurs and (ii) none of the multiple find
6 criteria are found to match the field content, and wherein the means for inserting the
7 replacement value comprises inserting the replacement value corresponding to one find
8 criteria that matched the field content.

1 26. The system of claim 14, wherein the rule definitions include a row clean
2 flag, and wherein at least one rule definition has the row clean flag set, further comprising
3 removing any row including a field matching the search criteria from the input table when
4 the row clean flag is set.

1 27. An article of manufacture for use performing a clean operation on an input
2 table in a database having an input data table name, the article of manufacture comprising
3 computer usable media including at least one computer program embedded therein that
4 causes the computer to perform:

5 receiving at least one rule definition, wherein each rule definition indicates a find
6 criteria, a replacement value, and an input data column in the input table;

7 searching, for each rule definition, the input data column for any fields that match
8 the find criteria; and

9 inserting, for each rule definition, the replacement value in the fields in the input
10 data column that match the find criteria, wherein subsequent applications of additional
11 rule definitions applied to the same input data column operate on replacement values
12 inserted in the input data column in previously applied rule definitions.

1 28. The article of manufacture of claim 27, wherein each rule definition is
2 associated with one rule table including the find criteria and replacement value, wherein a
3 rule table column parameter for each rule definition indicates the columns in the rule
4 table including the find criteria and replacement value for the rule definition.

1 29. The article of manufacture of claim 27, wherein there is a separate rule
2 table including the find criteria and replacement value associated with at least one rule
3 definition, wherein, for each rule definition, a rule table column parameter indicates the
4 columns in the rule table for the rule definition including the find criteria and replacement
5 value for that rule definition.

1 30. The article of manufacture of claim 27, wherein the input data column for
2 a first and second applied rule definitions is the same input data column, wherein the
3 replacement value for the first rule definition is inserted into at least one field in the input

1 data column, and wherein the find criteria of the second rule definition is applied to the
2 replacement value inserted in the input data column.

1 31. The article of manufacture of claim 27, wherein at least one rule definition
2 includes multiple find criteria and a corresponding replacement value for each find
3 criteria, wherein the step of searching the input data column comprises applying each of
4 the multiple find criteria to one field until one of: (i) a match occurs and (ii) none of the
5 multiple find criteria are found to match the field content, and wherein inserting the
6 replacement value comprises inserting the replacement value corresponding to one find
7 criteria that matched the field content.

1 32. The article of manufacture of claim 31, wherein a sort column includes
2 values to use to sort the multiple find criteria and corresponding replacement value,
3 wherein the step of searching comprises applying the multiple find criteria to each field in
4 the order specified in the sort column.

1 33. The article of manufacture of claim 27, wherein the rule definition
2 comprises a type of rule that is a member of the set of rules consisting of: find and
3 replace, discretization, and numeric clip, wherein at least two rule definitions are
4 comprised of different rule types.

1 34. The article of manufacture of claim 27, wherein the find criteria for at least
2 one rule definition comprises an upper bound and lower bound, wherein searching
3 comprises searching for any fields that have values within the upper and lower bounds.

1 35. The article of manufacture of claim 34, wherein the at least one rule
2 definition including find criteria having upper and lower bounds includes multiple find
3 criteria and a corresponding replacement value for each find criteria, wherein the step of

4 searching the input data column comprises applying each of the multiple find criteria to
5 one field until one of: (i) a match occurs and (ii) none of the multiple find criteria are
6 found to match the field content, and wherein inserting the replacement value comprises
7 inserting the replacement value corresponding to one find criteria that matched the field
8 content.

1 36. The article of manufacture of claim 34, wherein searching comprises
2 searching for any fields that have values outside of one of the upper and lower bounds.

1 37. The article of manufacture of claim 27, wherein the find criteria for at least
2 one rule definition comprises an upper bound and lower bound and wherein the
3 replacement value is an upper replacement value and further comprising a lower
4 replacement value, wherein searching comprises searching for any fields that have values
5 within the upper and lower bounds and wherein inserting comprises inserting the upper
6 replacement value if the field has a value greater than the upper bound and inserting the
7 lower replacement value if the field has a value less than the lower bound.

1 38. The article of manufacture of claim 37, wherein the at least one rule
2 definition including find criteria having upper and lower bounds includes multiple find
3 criteria and a corresponding upper and lower replacement value for each find criteria,
4 wherein the step of searching the input data column comprises applying each of the
5 multiple find criteria to one field until one of: (i) a match occurs and (ii) none of the
6 multiple find criteria are found to match the field content, and wherein inserting the
7 replacement value comprises inserting the replacement value corresponding to one find
8 criteria that matched the field content.

1 39. The article of manufacture of claim 27, wherein the rule definitions
2 include a row clean flag, and wherein at least one rule definition has the row clean flag

1 set further comprising removing any row including a field matching the search criteria
2 from the input table when the row clean flag is set.

1 40. A memory device including a command for performing a clean operation
2 on a computer database input table, the command comprising
3 an input data table name parameter indicating the input table subject to the clean
4 operation; and
5 at least one rule definition, wherein each rule definition includes a find criteria, a
6 replacement value, and an input data column in the input table, wherein, for each rule
7 definition, the input data column is searched for any fields that match the find criteria,
8 wherein, for each rule definition, the replacement value is inserted in the fields in the
9 input data column that match the find criteria, and wherein subsequent applications of
10 additional rule definitions applied to the same input data column operate on replacement
11 values inserted in the input data column in previously applied rule definitions.

1 41. The memory device of claim 40, wherein at least one rule definition
2 further includes:
3 indication of one rule table including the find criteria and replacement value for
4 the at least two rule definitions, such that the one rule table includes the find criteria and
5 replacement value for the at least two rule definitions; and
6 a rule table column parameter for the at least two rule definitions indicating the
7 columns in the rule table including the find criteria and replacement value for the rule
8 definitions.

1 42. The memory device of claim 40, wherein at least one rule definition
2 further includes:
3 indication of a separate rule table for each rule definition including the find
4 criteria and replacement value for the at least two rule definitions; and

1 a rule table column parameter indicating the columns in the rule table for the rule
2 definition including the find criteria and replacement value for that rule definition.

1 43. The memory device of claim 40, wherein the input data column for a first
2 and second applied rule definitions is the same input data column.

1 44. The memory device of claim 40, wherein at least one rule definition
2 further includes:
3 multiple find criteria and a corresponding replacement value for each find criteria,
4 wherein the input data column is searched by applying each of the multiple find criteria to
5 one field until one of: (i) a match occurs and (ii) none of the multiple find criteria are
6 found to match the field content, and wherein the replacement value corresponding to the
7 matching find criteria is inserted into the field.

1 45. The memory device of claim 4, wherein the at least one rule definition
2 further comprises a sort column including values to use to sort the multiple find criteria
3 and corresponding replacement value, wherein the multiple find criteria are applied to
4 each field in the input data column in the order specified in the sort column.

1 46. The memory device of claim 40, wherein the rule definition comprises a
2 type of rule that is a member of the set of rules consisting of: find and replace,
3 discretization, and numeric clip, wherein at least two rule definitions are comprised of
4 different rule types.